

A novel approach for detection of cracks in painting and concrete surface images using CNN models

Deepti Vadicherla¹, Poonam Gupta^{1,2}

¹Department of Computer Science and Engineering, GH Raisoni University, Amravati, India

²Department of Information Technology, GH Raisoni College of Engineering and Management, Pune, India

Article Info

Article history:

Received Dec 24, 2024

Revised Jul 31, 2025

Accepted Oct 15, 2025

Keywords:

Classification

CNN

Inception V3

Machine learning

ResNet-50

VGG-16

ABSTRACT

Discovering the beginnings of historical artworks takes one on an amazing voyage across space and time. People all around the world have been captivated by India's rich cultural heritage throughout its history, and ancient paintings have always been a very important part of it. Over the period of time, these ancient paintings can get cracks on it due to many factors. This research introduces an automated image classification system where the cracks on the paintings as well as the concrete surface will get detected. Detecting cracks on the concrete surface is important because the longevity and upkeep of concrete structures rely on the prompt identification and treatment of cracks, which can weaken the structure and necessitate expensive repairs. In this study, we focus on image classification using general convolution neural network (CNN), Inception V3, VGG-16, and ResNet-50 models of CNN. These models are trained and validated separately on two different datasets of paintings and concrete surfaces. Inception V3 and VGG-16 models achieve high accuracy, respectively in painting and concrete datasets in comparison with general CNN and ResNet-50 models.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Deepti vadicherla

Department of Computer Science and Engineering, GH Raisoni University

444701 Maharashtra, India

Email: deepti.vadicherla@gmail.com

1. INTRODUCTION

Every nation's creative past bears witness to the unbounded imagination, profound spiritual understanding, and rich cultural diversity that enthrall and inspire people throughout. An artwork's materials, hues, and symbolism might reveal details about the civilisation that produced it. We can see people, places, and events from a wider range of perspectives because to artwork. By deconstructing ancient artworks and analysing their details, we may experience a new age and journey back in time.

Old paintings are that which have already survived hundreds of years and that on both canvas and wood panels with their layers of oil. Although they can be damaged by many factors like, the damages caused by the environmental factors [1] such as exposure to sunlight and moisture, wrong restoration attempts can be damaging if they're done by unskilled people or if restorers use techniques that are later found to be harmful. Breaks in the paint, varnish, or substrate affect the ancient paintings. We were able to preserve the beauty of these paintings when we digitized them by identifying [2] and removing cracks [3] and restoring them with mathematical methods [4], [5].

For facilities to remain resilient, secure, and durable all over time, it is vital to locate the presence of cracks on its surfaces of concrete [6]. Also, for the purpose of averting possible threats and expensive fixes, early crack recognition is indispensable [7], [8]. These crevices have the capacity to compromise the

structural integrity of a structure, thereby threatening the security of its people. Proactive maintenance measures that enhance the durability of concrete and minimize expenses can be caused by early rupture identification [9]. Prompting interference reduces the requirement for broader, more costly restorations by hindering the harm from progressing. Crack detecting is not simply cost-effective; it helps preserve sustainable development by minimizing the negative impacts that construction breakdowns constitute to the surroundings [10]. It is also necessary for protecting asset management, establishing quality control during the entire construction process, and improving security for the public, visual impact, and overall confidence in the reliability of concrete buildings. The manual way of identification of cracks in old images has some limitations. This paper proposes a study to find the efficient approach to identify cracks in images using machine learning techniques.

Machine learning, which includes many algorithms for effectively developing their version on a certain topic, has emerged as the most influential tool [11], [12]. It is often acknowledged that machine learning possesses remarkable abilities in classifying images and identifying patterns in data sets [13]. Compared to previous methods that rely on rule-based systems and manual feature extraction, machine learning [14] offers substantial advantages in picture classification, such as automation, high accuracy, and the capacity to handle massive datasets.

This article provides an approach for image classification of ancient paintings as well as concrete surface with cracks. Many scholarly articles were studied from journals and elite conferences and the image classification methods [15], [16] of machine learning were compared. For image classification, generic convolution neural network (CNN), Inception V3, VGG-16, and ResNet-50 are selected among these four techniques. This paper is structured as below. Chapter 2 gives the insight about the general CNN architecture and the sources of datasets and discuss about need of image augmentation technique and the discuss different augmentation functions used for painting image dataset to increase its size. Based on the analysis of many reference works, three CNN models Inception V3, VGG-16, and ResNet-50 are discussed. These four CNN methods were trained and validated using two datasets and its results are discussed in chapter 3. Chapter 4 concludes the study.

2. METHOD

The goal of this paper to analyze the crack detection process on two datasets using different tools in machine learning techniques. CNN are one type of DL method that is widely utilised for image organization and other object identification applications. CNN performs well in image classification [17]. In this paper, we are using two datasets with different surfaces, one is a painting dataset and other one is the cement surface dataset. The CNN model is contrasted with different CNN models, such as Inception V3, VGG-16 and ResNet-50.

2.1. CNN structure

This popular design is the CNN; a supervised ML method mostly used for image and voice recognition [18]. One regularised type of feed-forward neural network that learns features on its own is a convolutional neural network, which uses filter or kernel optimisation. CNN learns to optimise the filters by automated learning, as opposed to traditional algorithms that hand-engineer the filters [19]. The fact that feature extraction is independent of prior knowledge and human interaction is a major advantage. Figure 1 depicts CNN's architecture.

A simple CNN model consists of an input layer, hidden layers, and an output layer. A CNN's hidden layers contain one or more convolution-performing layers [20]. This method consists of a layer that performs the dot product using the convolution kernel and the layer's input matrix. As the convolution kernel traverses the layer's input matrix, the convolution process generates a feature map, which is subsequently fed into the layer's input [21]. This is followed by further layers such as normalisation layers, pooling layers, and fully linked layers.

Convolution and pooling are the two primary processes of a CNN classifier. Convolutional layers forward the output of the input's convolution to the subsequent layer. To speed up processing, depth-wise separable convolutional layers can be employed in place of standard convolutional layers. Convolutional networks can incorporate local or global pooling layers in addition to the standard convolutional layers. Pooling layers reduce the dimensionality of data by combining the outputs of neurone clusters in one layer into a single neurone in the next layer [22], [23]. Through fully linked layers, every neurone in one layer may communicate with every other layer's neurone. It goes through many activation layers to get desired final output. In this paper we have used CNN classifier models to obtain an effective classification of our painting and concrete surface images.

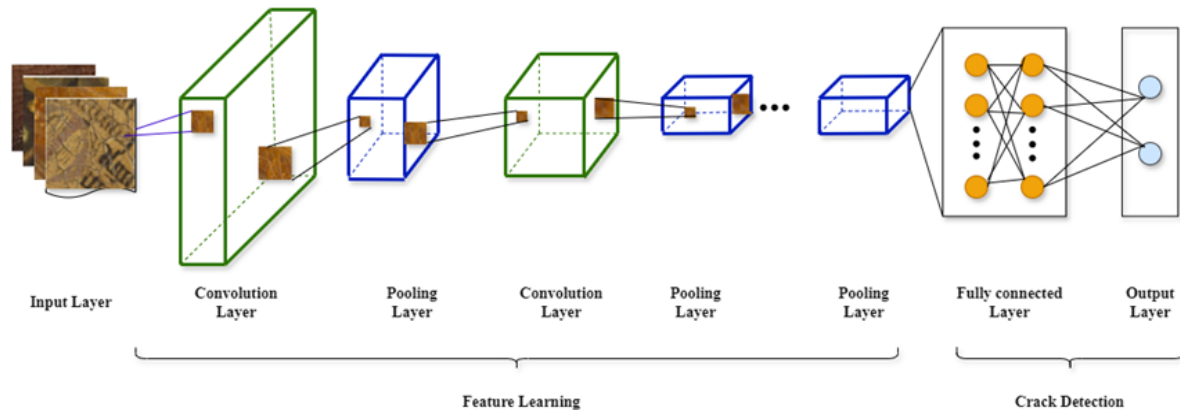


Figure 1. A general idea of a CNN-based classifier

2.2. Data collection

To empirically test the effectiveness of the suggested strategy, two datasets of digital photographs of paintings and concrete images are gathered from publicly accessible internet repositories. In case of painting dataset different CNN methods were tested earlier on a small dataset of about 780 images which later increased to 2,060 images using augmentation technique.

2.2.1. Image augmentation technique

In order to increase the accuracy and versatility of machine learning models, especially for the image classification, researchers frequently employ data augmentation approaches [24]. We have employed the data augmentation technique to expand the painting dataset in order to improve image classification results [25]. The augmentation functions are applied using the ImageDataGenerator class from TensorFlow's Keras module as shown below.

```
datagen = ImageDataGenerator(
    rotation_range=40,
    width_shift_range=0.2,
    height_shift_range=0.2,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    fill_mode='nearest'
)
```

Augmentation functions artificially expand and diversity datasets by applying changes to preexisting data. `rotation_range=40` is the rotation function, rotates the picture up to 40 degrees at random. The `width_shift_range=0.2` function is used to randomly shift the picture horizontally by up to 20% of its width. The picture is randomly shifted vertically by up to 20% of its height when `height_shift_range=0.2` is used. A shear transformation (tilting) of up to 20% is applied to the picture by function `shear_range=0.2`. The picture may be arbitrarily zoomed in or out by up to 20% with `zoom_range=0.2`. To flip the image horizontally we used the function `horizontal_flip=True`. After transformations, the `fill_mode='nearest'` function uses the values of the closest nearby pixels to fill in newly produced pixels.

Following the augmentation procedure, we obtained a sizable painting dataset with 2,060 images. The concrete surface dataset which we have used from Kaggle website has 5,000 images were sufficient for the image classification, so we have applied the augmentation technique only to the painting dataset.

2.2.2. Painting image dataset

Painting image classification data is collected by gathering photos from different resources, databases, and museum repositories which are available online. Figure 2 shows the painting images with cracks which are obtained from the ghent altarpiece website [26]. The early Netherlandish artists Hubert and Jan van Eyck are credited with creating the enormous, elaborate polyptych altarpiece from the 15th century that is kept at St. Bavo's Cathedral in Ghent, Belgium. These images which are included in the dataset, are of very high resolution taken after restoration stage with macrophotography modality. Knights of christ [27]

panel from the Ghent Altarpiece, which measures approximately 51.4 cm in width and 146.2 cm in height as shown in Figure 2(a), is mainly used to add crack painting images in our dataset. Crack images are cropped and are used in the scale 545 pixels/cm shown in Figure 2(b).

Machine learning models are trained with image classification datasets to enable them to identify and categorize images into predetermined groups. Each training image has been labelled manually, classifying each one according to its content and assigning it to the category of images with cracks and without cracks. Initially we got a set of cropped painting images of both classes, then we used augmentation method to increase the size of our painting dataset. The total 2,060 images has been obtained after using augmentation technique, which later used in classification process. Among the total images, 1,440 images used as training dataset and 620 images are used in testing process. Painting images with and without cracks are shown in Figure 3.



Figure 2. Painting with cracks (a) original painting images and (b) cropped image



Figure 3. Examples of different crack appearances in paintings

2.2.3. Concrete surface dataset

The dataset consists of concrete surface images, it is divided into two categories as with cracks and without cracks. The data is collected from Kaggle website which is a machine learning community with large resources [28]. This dataset includes 5,000 images with balanced classes of images with and without cracks. There was a lot of variation in the surface polish and lighting condition of these high-resolution photographs. These images are with 227×227 pixels dimensions. Concrete surface images with and without cracks are shown in Figure 4.

Ensuring that every feature of the image has an equivalent range, it is crucial to keeping the gradients from spiralling out of control. Since these CNNs leverage the idea of parameter sharing, values sharing will be difficult if the network inputs are not scaled to have comparable known ranges. This is because various regions of the image may end up containing values from completely different domains and ranges.



Figure 4. Sample images of concrete surface

2.3. CNN models

This study employed three convolutional neural network models along with general CNN model to do the image classification of surfaces of painting and concrete with and without cracks. Inception V3, VGG 16, and ResNet-50 are the three models of CNN.

2.3.1. Inception V3 model

An advanced CNN model with 48 layers, Inception V3 is capable of recognising and learning complex patterns and attributes from pictures. This third edition of Google's Inception CNN was designed to support deeper networks while preventing an excessive number of parameters from being used [29].

One of Inception V3's main characteristics is its ability to scale to massive datasets and handle pictures of various sizes and resolutions. Inception-v3 has receptive fields of different widths because it uses convolutional kernels of different sizes. To reduce the design space and accomplish feature fusion at various sizes, the network employs a modular structure with final connecting [30]. Large convolution kernels in Inception-v3 are split up into smaller convolution kernels in sequence, and pooling and convolution are coupled in parallel.

In the proposed research the dataset of painting images is used to train and test the Inception v3 model, during this process we have studied multiple reputed research papers. Authors used the malware signature image dataset to compare the performance of technologies linear regression (LR), artificial neural network (ANN), CNN. InceptionV3's transfer learning strategy produced a strong performance [31]. The Inception V3 model is used as a foundation in another study, and a fully linked layer is constructed on top of it to maximise the classification process. Prior to the classification stage, the custom model is concatenated with the independently acquired segmented features. It improved the functionality of key characteristics and was used to the food categorisation process [32]. The method used here is based on using neural networks for classification and Inception V3 for feature extraction. Six different sports have been examined and categorised. Over six categories, the framework's average accuracy was 96.64% [33]. Inception V3 model presents a promising tool which works better for concrete surface dataset [34] and beyond that improving industrial quality assurance in steel manufacturing [35].

2.3.2. VGG-16

The VGG-16 model is a CNN architecture proposed by the visual geometry group (VGG) at the University of Oxford. It stands out for its depth with a total of 16 layers, of which 13 are convolutional layers and 3 are fully linked layers. For picture recognition and classification, VGG-16 is often utilised [36], [37]. It is a straightforward and efficient technique for classifying images.

Max-pooling layers come after a stack of convolutional layers that become deeper in the model's architecture. The model's architecture enables it to learn intricate hierarchical representations of visual data, producing precise and dependable predictions. Even though VGG-16 is less complex than more modern architectures, it is still a popular choice for many deep learning applications because of its outstanding speed and adaptability [38].

The painting image dataset used in Inception v3 classification is also used to train and test VGG-16 model. Acquired results are discussed in next section. Different research papers have been studied to analyse results based on other datasets. Ye [39] suggested a lightweight model based on VGG-16 that can identify and categorise remote sensing pictures, eliminate duplicate information, and selectively extract certain characteristics. In addition to guaranteeing correctness, this model lowers the model's parameters.

Bhosale [40] investigated the effectiveness of the VGG-16 and ResNet-34 algorithms of the CNN for supervised classification-based land use land cover change (LULCC) identification. Medical field related dataset is used for image classification using VGG-16 model. A VGG stacked classifier network was

suggested by the authors in [41] to classify brain tumour pictures with 99% accuracy without the need for human involvement. Three art classification datasets were used in the studies [42] and the results showed that the suggested approach significantly outperforms the current baseline methods, where suggested approach was evaluated with a shallow neural network serving as the second-stage classifier and six distinct pre-trained CNNs serving as the first-stage classifiers.

2.3.3. ResNet-50

ResNet is a model for classifying images. In addition to one MaxPool and one average pool layer, a ResNet model variant known as ResNet50 has 48 convolution layers [43]. ResNet created the residual learning block as a solution to the disappearing gradient problem. An increase in accuracy was observed in ResNet50 when skip connections were employed to maintain the gradient in the deeper layer [44].

The key drivers of innovation come from residual blocks, which adopt a congestion design of 1x1, 3x3, and 1x1 convolutions that are significant source of innovation. To lessen the effect of the vanishing gradient, remaining connections also referred to as skip connections are used. The product of one layer can be added instantly to the next layer thanks to these linkages. Rather than employing the traditional fully connected layers, global average pooling reduces the geographical dimensions to a 1x1 grid [45].

A fully connected layer that classifies information via softmax stimulation completes its construction. ResNet-50 is a standard that has demonstrated efficacy in image processing problems since it can help with the research of deep networks by utilizing the remaining connections. Work focused on resnet50 model which used fire images to study the algorithms [46]. The findings demonstrated that, with an accuracy rate of 98%, the resnet50 model is more accurate than the suggested CNN algorithm. Authors combined two mainstream computer vision models, Resnet 3D and video swin transformer [47]. The proposed model is compared with ResNet-50. A ResNet-50-based diagnostic method for precise breast abnormality categorisation is introduced [48]. Using the ISIC archive dataset [49], this study presents an automated skin cancer classification system with an emphasis on deep learning. Dermoscopic image classification is done using the VGG-16 and ResNet-50 models. For feature extraction, the work integrates a spatial pyramid pooling layer into a modified ResNet-50 architecture with transfer learning.

3. RESULTS AND DISCUSSION

This section presents the findings and analysis of our suggested model, demonstrating the classification process's performance evaluation of two dataset photos of painted and cracked concrete surfaces. Models are developed using Google Colab platform with Python version 3.13 and the experimental setup done for this is system with Intel Core processor i5-6500U CPU 2.40GHz, our system was configured with 8GB RAM, windows 11-64bit operating system.

3.1. Experimental result

Inception V3, VGG-16, and ResNet-50 models were used to classify 2060 pictures from the painting image collection. Two basic metrics accuracy and loss are used to assess how well these classification models perform during training and validation. In Figures 5-8, the accuracy of the general CNN model can be seen in Figure 5(a), Inception V3 in Figure 6(a), VGG-16 in Figure 7(a), and ResNet-50 in Figure 8(a). In those graphs the curve in blue shows the accuracy of the model during training phase whereas the orange-coloured curve shows the accuracy during validation phase. The loss curves for the generic CNN, Inception V3, VGG-16, and ResNet-50 models are displayed in Figures 5(b), 6(b), 7(b), and 8(b) accordingly. The curve colours used match those of the accuracy curve. The loss value determined during each epoch is shown on these loss curve.

One whole run through the network of the entire training dataset is represented by an epoch. Measuring how many times the model has viewed the complete dataset during training is the basic idea. In line graph, X axis shows the accuracy or loss in respective figure and Y axis shows the values of epoch. Lines connect the data points that are plotted on the axes. The graph shows how the dependent variable changes with any deviations in the independent variables.

In the concrete surface dataset, 5000 images have been used in the classification process. In Figures 9-12, The accuracy of the generic CNN model is shown in Figure 9(a), Inception V3 in Figure 10(a), VGG-16 in Figure 11(a), and ResNet-50 in Figure 12(a) during the training and validation phases. These graphs indicate how accuracy varies with epoch. Figures 9(b), 10(b), 11(b), and 12(b) display the loss curves for the generic CNN, Inception V3, VGG-16, and ResNet-50 algorithms, respectively, using the same dataset.

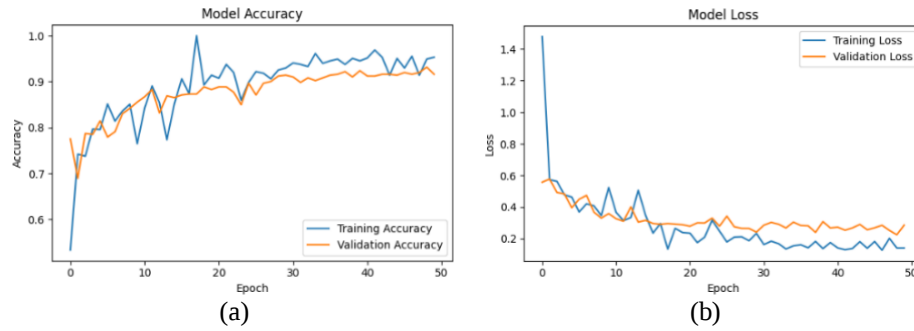


Figure 5. Experimental results of general CNN model using painting image dataset
(a) accuracy curve and (b) loss curve

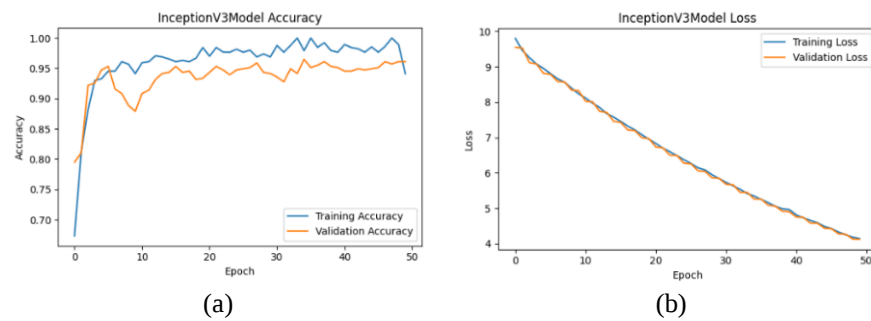


Figure 6. Experimental results of Inception v3 model using painting image dataset
(a) accuracy curve and (b) loss curve

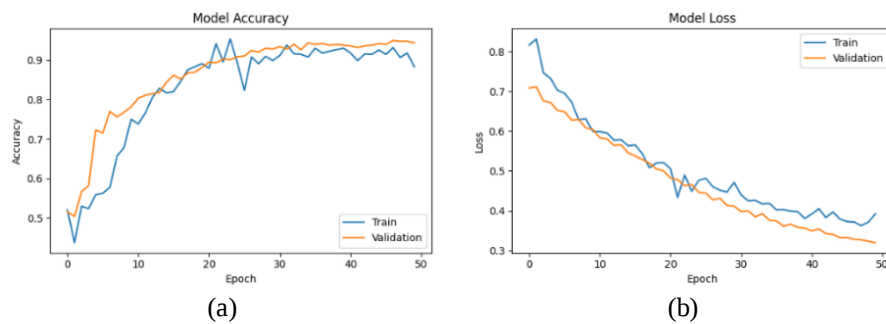


Figure 7. Experimental results of VGG-16 model using painting image dataset
(a) accuracy curve and (b) loss curve

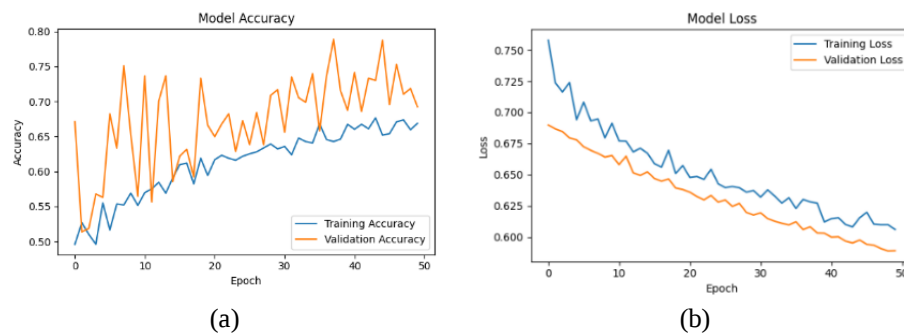


Figure 8. Experimental results of ResNet-50 model using painting image dataset
(a) accuracy curve and (b) loss curve

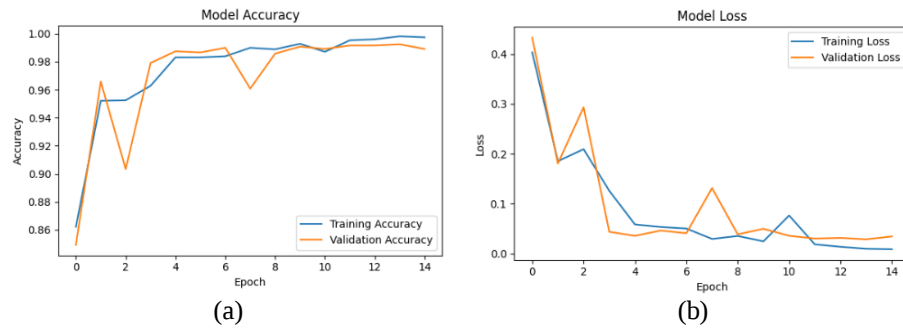


Figure 9. Experimental results of general CNN model for concrete surface dataset
(a) accuracy curve and (b) loss curve

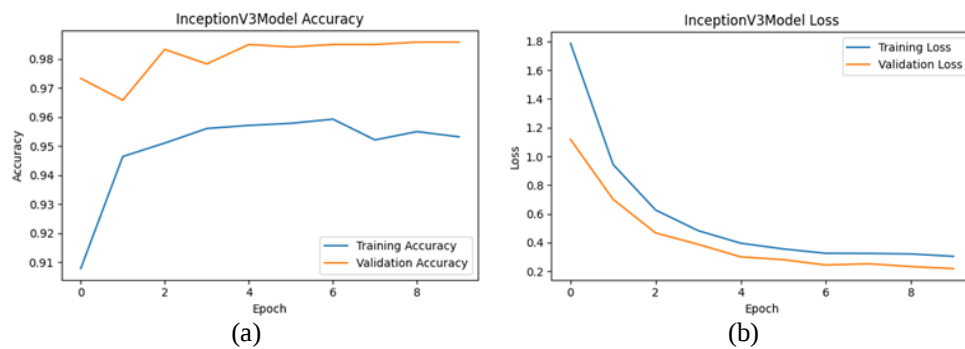


Figure 10. Experimental results of Inception V3 model for concrete surface dataset

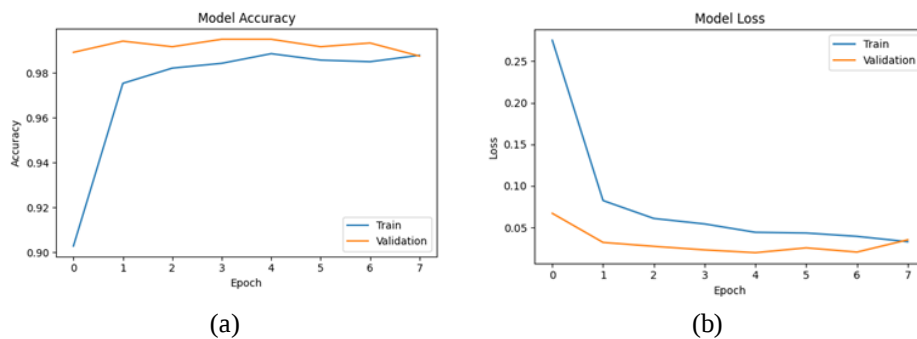


Figure 11. Experimental results of VGG-16 model using concrete surface dataset
(a) accuracy curve and (b) loss curve

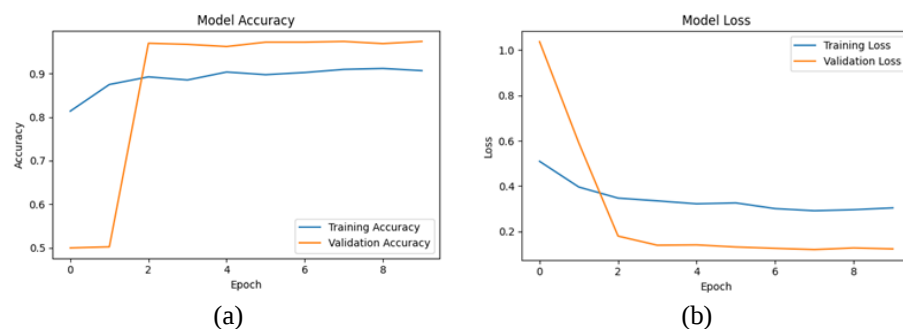


Figure 12. Experimental results of ResNet-50 model using concrete surface dataset
(a) accuracy curve and (b) loss curve

3.2. Model evaluation

This section presents the experimental findings of four CNN models utilising a variety of assessment criteria, such as sensitivity, specificity, recall, accuracy, and precision [50], [51]. An approach that is frequently used to evaluate a trained model's prediction power in relation to a certain validation dataset is the confusion matrix. In Figure 13, display the confusion matrices for the generic CNN is shown in Figure 13(a), Inception V3 Figure 13(b), VGG-16 Figure 13(c), and ResNet-50 Figure 13(d) algorithms using the painting picture dataset, respectively. In Figure 14, the confusion matrices for the generic CNN, Inception V3, VGG-16, and ResNet-50 models utilising specific picture datasets are shown in Figures 14(a), 14(b), 14(c), and 14(d). The purpose of the confusion matrix is to show how accurate or confused the model is. True positive (tp), true negative (tn), false positive (fp), and false negative (fn) class values can be obtained from these matrices. True positives indicate the number of properly identified positive samples, whereas true negatives indicate the number of correctly predicted negative samples.

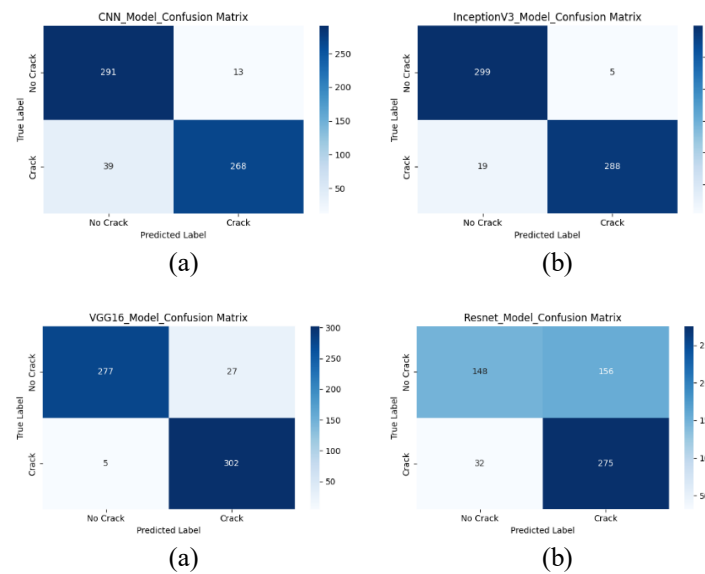


Figure 13. For Painting image dataset confusion matrices of models, (a) general CNN, (b) inception V3, (c) VGG-16, and (d) ResNet-50

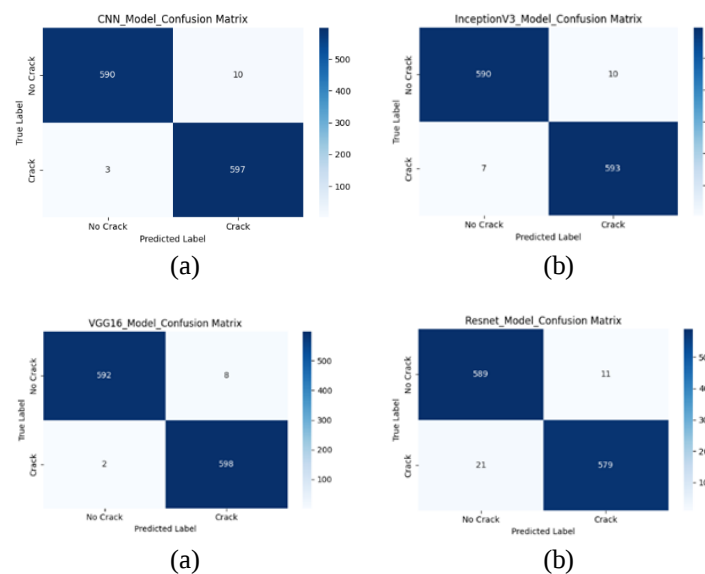


Figure 14. For concrete image dataset confusion matrices of models, (a) general CNN, (b) inception V3, (c) VGG-16, and (d) ResNet-50

The values that are shown in the Tables 1 and Table 2 give information about the performance of all four CNN models tried on two different datasets. The values in the tables are carefully calculated using a series of particular mathematical formulas as given in the (1)-(4), providing important information on the forecasting ability of the model. Figure 15 shows the accuracy comparison of all four models in the chart form. In this comparison it is observed that, Inception V3 model achieved high accuracy when we trained all four models with painting image dataset, while in case of concrete surface dataset VGG-16 model shown the slightly greater accuracy than that of accuracy achieved by general CNN, InceptionV3 and ResNet-50 models.

$$\text{Accuracy} = \frac{tp+tn}{tp+tn+fp+fn} \quad (1)$$

$$\text{Precision} = \frac{tp}{tp+fp} \quad (2)$$

$$\text{Sensitivity} = \frac{tp}{tp+fn} \quad (3)$$

$$\text{Specificity} = \frac{tn}{tn+fp} \quad (4)$$

Table 1. Comparative analysis of four CNN models using painting image dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	Sensitivity (%)	Specificity (%)
CNN	91.49	95.37	87.30	87.30	95.72
Inception V3	96.07	98.29	93.81	93.81	98.36
VGG16	94.76	91.79	98.37	98.37	91.12
ResNet-50	69.23	63.81	89.58	89.58	48.68

Table 2. Comparative analysis of four CNN models using concrete surface dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	Sensitivity (%)	Specificity (%)
CNN	92.95	93.51	92.31	92.31	93.59
Inception V3	87.82	85.54	91.03	91.03	84.62
VGG16	94.87	94.87	94.22	94.87	94.87
ResNet-50	92.31	93.42	91.03	91.03	93.59

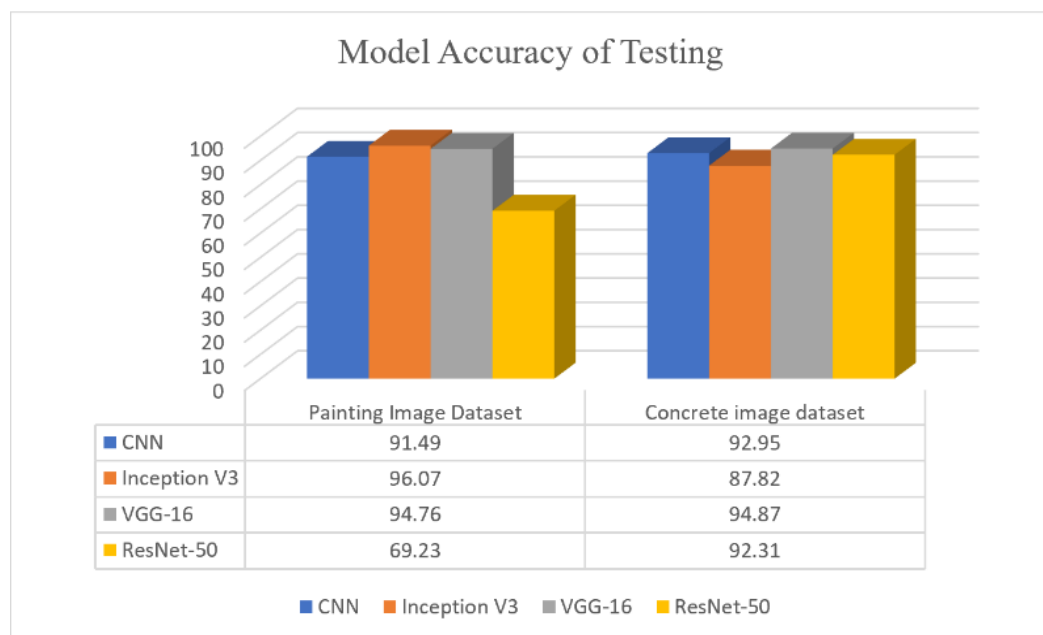


Figure 15. Testing accuracy of each selected model

4. CONCLUSION

We have implemented a remedy in this work to address the issue of detecting the cracks in painting images and concrete surfaces, in future which can be further treated with restoration process to remove the artifacts present in the painting images and in case of concrete surface cracks to enumerate precautionary measures to avoid further damage. This study looked into comparing four different strategies, general CNN, Inception V3, VGG-16 and ResNet-50. These four different models were applied on the two datasets to achieve the high accuracy in classification result

Four CNN models are trained and tested using both datasets. The evaluation is done in terms of accuracy. The findings revealed that, in case of painting dataset the accuracy achieved by Inception V3 is greater than that of general CNN, VGG-16 and ResNet-50 models. Regarding concrete surface dataset, VGG-16 model give good accuracy than other three models. These results suggest that VGG-16 model achieves high accuracy and it is effective to classify cracks on painting and concrete surfaces.

The study demonstrates the performance gap in the current models with an emphasis on enhancing the system's accuracy, assisting in the restoration of ancient art and assisting in the early detection of damage. However, recognizing the many limitations imposed by current models is crucial. The thin hair like structure in paintings can be recognised as crack on the surface. To train the system with such images it becomes difficult to add painting images without cracks but the part of painting which looks like crack. Due to limitations of datasets the models can not give more accurate results than the current obtained results. However, further the classification approach can be extended in future with detecting multiple types of cracks in the more complicated form in painting as well as on the monuments.

FUNDING INFORMATION

Authors state no funding involved.

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.

REFERENCES

- [1] I. Giakoumis, N. Nikolaidis, and I. Pitas, "Digital image processing techniques for the detection and removal of cracks in digitized paintings," *IEEE Transactions on Image Processing*, vol. 15, no. 1, pp. 178–188, Jan. 2006, doi: 10.1109/TIP.2005.860311.
- [2] R. Sizyakin *et al.*, "Crack detection in paintings using convolutional neural networks," *IEEE Access*, vol. 8, pp. 74535–74552, 2020, doi: 10.1109/ACCESS.2020.2988856.
- [3] B. Cornelis *et al.*, "Crack detection and inpainting for virtual restoration of paintings: The case of the Ghent Altarpiece," *Signal Processing*, vol. 93, no. 3, pp. 605–619, Mar. 2013, doi: 10.1016/j.sigpro.2012.07.022.
- [4] Y. Zeng and Y. Gong, "Nearest neighbor based digital restoration of damaged ancient chinese paintings," in *International Conference on Digital Signal Processing, DSP*, IEEE, Nov. 2018, pp. 1–5. doi: 10.1109/ICDSP.2018.8631553.
- [5] J. Yang, W. Wang, G. Lin, Q. Li, Y. Sun, and Y. Sun, "Infrared thermal imaging-based crack detection using deep learning," *IEEE Access*, vol. 7, pp. 182060–182077, 2019, doi: 10.1109/ACCESS.2019.2958264.
- [6] B. Li, H. Guo, Z. Wang, and F. Wang, "Automatic concrete crack identification based on lightweight embedded U-Net," *IEEE Access*, vol. 12, pp. 148387–148404, 2024, doi: 10.1109/ACCESS.2024.3416450.
- [7] H. N. Nguyen, T. Y. Kam, and P. Y. Cheng, "Automatic crack detection from 2D images using a crack measure-based B-spline level set model," *Multidimensional Systems and Signal Processing*, vol. 29, no. 1, pp. 213–244, Jan. 2018, doi: 10.1007/s11045-016-0461-9.
- [8] A. S. Rao, T. Nguyen, M. Palaniswami, and T. Ngo, "Vision-based automated crack detection using convolutional neural networks for condition assessment of infrastructure," *Structural Health Monitoring*, vol. 20, no. 4, pp. 2124–2142, Jul. 2021, doi: 10.1177/1475921720965445.
- [9] C. Shao, L. Zhang, and W. Pan, "PTZ camera-based image processing for automatic crack size measurement in expressways," *IEEE Sensors Journal*, vol. 21, no. 20, pp. 23352–23361, Oct. 2021, doi: 10.1109/JSEN.2021.3112005.
- [10] H. Lang, Y. Yuan, J. Chen, S. Ding, J. J. Lu, and Y. Zhang, "Augmented concrete crack segmentation: learning complete representation to defend background interference in concrete pavements," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–13, 2024, doi: 10.1109/TIM.2024.3378205.
- [11] V. Chaudhari, A. Bajpai, S. Amrutkar, S. Dubey, and D. P. Gupta, "Introduction to machine learning and its application," *International Journal for Research in Applied Science and Engineering Technology*, vol. 11, no. 1, pp. 1749–1754, Jan. 2023, doi: 10.22214/ijraset.2023.48899.
- [12] R. Praba, G. Darshan, T. K. Roshanraj, and B. S. Prakash, "Study on machine learning algorithms," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, pp. 67–72, Jul. 2021, doi: 10.32628/cseit2173105.

- [13] J. Li, P. Chen, S. Yu, S. Liu, and J. Jia, "MOODv2: Masked image modeling for out-of-distribution detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 46, no. 12, pp. 8994–9003, Dec. 2024, doi: 10.1109/TPAMI.2024.3412004.
- [14] S. A. A. Kharis, A. H. A. Zili, M. Malik, W. Nuryaningrum, and A. Putri, "Comparing machine learning models for Indonesia stock market prediction," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 38, no. 1, p. 508, 2025, doi: 10.11591/ijeecs.v38.i1.pp508-516.
- [15] H. C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN Architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, May 2016, doi: 10.1109/TMI.2016.2528162.
- [16] C. Broni-Bediako, Y. Murata, L. H. B. Mormille, and M. Atsumi, "Searching for CNN architectures for remote sensing scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022, doi: 10.1109/TGRS.2021.3097938.
- [17] R. Ashraf *et al.*, "Deep convolution neural network for big data medical image classification," *IEEE Access*, vol. 8, pp. 105659–105670, 2020, doi: 10.1109/ACCESS.2020.2998808.
- [18] L. Xu, S. Lv, Y. Deng, and X. Li, "A weakly supervised surface defect detection based on convolutional neural network," *IEEE Access*, vol. 8, pp. 42285–42296, 2020, doi: 10.1109/ACCESS.2020.2977821.
- [19] F. C. Chen and M. R. Jahanshahi, "NB-CNN: Deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4392–4400, May 2018, doi: 10.1109/TIE.2017.2764844.
- [20] L. Alzubaidi *et al.*, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.
- [21] J. A. AYENI, "Convolutional neural network (CNN): the architecture and applications," *Applied Journal of Physical Science*, vol. 4, no. 4, pp. 42–50, Dec. 2022, doi: 10.31248/ajps2022.085.
- [22] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, "Conceptual understanding of convolutional neural network- a deep learning approach," *Procedia Computer Science*, vol. 132, pp. 679–688, 2018, doi: 10.1016/j.procs.2018.05.069.
- [23] Y. Lee, "The CNN: the architecture behind artificial intelligence development," *Journal of Student Research*, vol. 12, no. 4, Nov. 2023, doi: 10.47611/jsrhs.v12i4.5579.
- [24] A. Mikołajczyk and M. Grochowski, "Data augmentation for improving deep learning in image classification problem," in *2018 International Interdisciplinary PhD Workshop, IIPHDW 2018*, IEEE, May 2018, pp. 117–122. doi: 10.1109/IIPHDW.2018.8388338.
- [25] C. Shorten and T. M. Khoshgoufar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, Dec. 2019, doi: 10.1186/s40537-019-0197-0.
- [26] Closer to Van Eyck, "The ghent altarpiece," [clostertovaneyck.kikirpa.be](https://clostertovaneyck.kikirpa.be/ghentaltarpiece/). Accessed: May. 19, 2025. [Online]. Available: <https://clostertovaneyck.kikirpa.be/ghentaltarpiece/>
- [27] Closer to Van Eyck, "Knights of christ," [clostertovaneyck.kikirpa.be](https://clostertovaneyck.kikirpa.be/ghentaltarpiece/#viewer/rep1=0&id1=5dbe2336ab876d1c9766ddbc1268e843). Accessed: May. 19, 2025. [Online]. Available: <https://clostertovaneyck.kikirpa.be/ghentaltarpiece/#viewer/rep1=0&id1=5dbe2336ab876d1c9766ddbc1268e843>
- [28] kaggle, "Datasets," [kaggle.com/datasets](https://www.kaggle.com/datasets). Accessed: May. 19, 2025. [Online]. Available: <https://www.kaggle.com/datasets>
- [29] Y. Pan *et al.*, "Fundus image classification using inception V3 and ResNet-50 for the early diagnostics of fundus diseases," *Frontiers in Physiology*, vol. 14, Feb. 2023, doi: 10.3389/fphys.2023.1126780.
- [30] J. Cao, M. Yan, Y. Jia, X. Tian, and Z. Zhang, "Application of a modified Inception-v3 model in the dynasty-based classification of ancient murals," *Eurasip Journal on Advances in Signal Processing*, vol. 2021, no. 1, p. 49, Dec. 2021, doi: 10.1186/s13634-021-00740-8.
- [31] M. Ahmed, N. Afreen, M. Ahmed, M. Sameer, and J. Ahamed, "An inception V3 approach for malware classification using machine learning and transfer learning," *International Journal of Intelligent Networks*, vol. 4, pp. 11–18, 2023, doi: 10.1016/j.ijin.2022.11.005.
- [32] V. C. Burkapalli and P. C. Patil, "Transfer learning: inception-V3 based custom classification approach for food images," *ICTACT Journal on Image and Video Processing*, vol. 11, no. 1, pp. 2245–2253, Aug. 2020, doi: 10.21917/ijivp.2020.0321.
- [33] K. Joshi, V. Tripathi, C. Bose, and C. Bhardwaj, "Robust sports image classification using inceptionv3 and neural networks," *Procedia Computer Science*, vol. 167, pp. 2374–2381, 2020, doi: 10.1016/j.procs.2020.03.290.
- [34] A. Hussain, K. N. Qureshi, R. W. Anwar, and A. Aslam, "A novel SCD11 CNN model performance evaluation with inception V3, VGG16 and ResNet50 using surface crack dataset," in *2nd International Conference on Unmanned Vehicle Systems-Oman, UVS 2024*, IEEE, Feb. 2024, pp. 1–7. doi: 10.1109/UVS59630.2024.10467149.
- [35] G. K. Uday, S. Gajula, A. Sravanthi, V. Srilakshmi, T. Vedithe, and D. V. Reddy, "Inception V3 model-based approach for detecting defects on steel surfaces," in *2nd International Conference on Intelligent Data Communication Technologies and Internet of Things, IDCIoT 2024*, IEEE, Jan. 2024, pp. 331–335. doi: 10.1109/IDCIOT59759.2024.10467368.
- [36] S. Tammina, "Transfer learning using VGG-16 with deep convolutional neural network for classifying images," *International Journal of Scientific and Research Publications (IJSRP)*, vol. 9, no. 10, p. p9420, Oct. 2019, doi: 10.29322/ijssrp.9.10.2019.p9420.
- [37] J. Tao, Y. Gu, J. Z. Sun, Y. Bie, and H. Wang, "Research on vgg16 convolutional neural network feature classification algorithm based on Transfer Learning," in *CISS 2021 - 2nd China International SAR Symposium*, IEEE, Nov. 2021, pp. 1–3. doi: 10.23919/CISS51089.2021.9652277.
- [38] Z. P. Jiang, Y. Y. Liu, Z. E. Shao, and K. W. Huang, "An improved VGG16 model for pneumonia image classification," *Applied Sciences (Switzerland)*, vol. 11, no. 23, p. 11185, Nov. 2021, doi: 10.3390/app112311185.
- [39] M. Ye *et al.*, "A lightweight model of VGG-16 for remote sensing image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 6916–6922, 2021, doi: 10.1109/JSTARS.2021.3090085.
- [40] V. Bhosale and A. Patankar, "Performance comparison of VGG-16 and ResNet-34 algorithms for supervised classification of landsat images," in *International Geoscience and Remote Sensing Symposium (IGARSS)*, IEEE, Jul. 2023, pp. 6247–6250. doi: 10.1109/IGARSS52108.2023.10281537.
- [41] M. S. Majib, M. D. M. Rahman, T. M. S. Sazzad, N. I. Khan, and S. K. Dey, "VGG-SCNet: A VGG net-based deep learning framework for brain tumor detection on MRI images," *IEEE Access*, vol. 9, pp. 116942–116952, 2021, doi: 10.1109/ACCESS.2021.3105874.
- [42] C. Sandoval, E. Pirogova, and M. Lech, "Two-stage deep learning approach to the classification of fine-art paintings," *IEEE Access*, vol. 7, pp. 41770–41781, 2019, doi: 10.1109/ACCESS.2019.2907986.
- [43] H. A. Khan and S. M. Adnan, "A novel fusion technique for early detection of alopecia areata using ResNet-50 and CRSHOG," *IEEE Access*, vol. 12, pp. 139912–139922, 2024, doi: 10.1109/ACCESS.2024.3461324.

- [44] M. B. Hossain, S. M. H. S. Iqbal, M. M. Islam, M. N. Akhtar, and I. H. Sarker, "Transfer learning with fine-tuned deep CNN ResNet50 model for classifying COVID-19 from chest X-ray images," *Informatics in Medicine Unlocked*, vol. 30, p. 100916, 2022, doi: 10.1016/j.imu.2022.100916.
- [45] A. Rath, B. S. P. Mishra, and D. K. Bagal, "ResNet50-based deep learning model for accurate brain tumor detection in MRI scans," *Next Research*, vol. 2, no. 1, p. 100104, Mar. 2025, doi: 10.1016/j.nexres.2024.100104.
- [46] H. Jabnoui, I. Arfaoui, M. A. Cherni, M. Bouhouicha, and M. Sayadi, "ResNet-50 based fire and smoke images classification," in *International Conference on Advanced Technologies for Signal and Image Processing, ATSIP 2022*, IEEE, May 2022, pp. 1–6, doi: 10.1109/ATSIP55956.2022.9805875.
- [47] C. Liu and C. Sun, "A fusion deep learning model of ResNet and vision transformer for 3D CT images," *IEEE Access*, vol. 12, pp. 93389–93397, 2024, doi: 10.1109/ACCESS.2024.3423689.
- [48] X. Yu, C. Kang, D. S. Guttery, S. Kadry, Y. Chen, and Y. D. Zhang, "ResNet-SCDA-50 for breast abnormality classification," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 94–102, Jan. 2021, doi: 10.1109/TCBB.2020.2986544.
- [49] S. Mejri and A. E. Oueslati, "Dermoscopic images classification using pretrained VGG-16 and ResNet-50 models," in *7th IEEE International Conference on Advanced Technologies, Signal and Image Processing, ATSIP 2024*, IEEE, Jul. 2024, pp. 342–347, doi: 10.1109/ATSIP62566.2024.10638943.
- [50] S. R. Shah, S. Qadri, H. Bibi, S. M. W. Shah, M. I. Sharif, and F. Marinello, "Comparing inception V3, VGG 16, VGG 19, CNN, and ResNet 50: a case study on early detection of a rice disease," *Agronomy*, vol. 13, no. 6, p. 1633, Jun. 2023, doi: 10.3390/agronomy13061633.
- [51] K. Neamah, F. Mohamed, S. R. Waheed, W. H. Madhloom, A. Y. Taha, and K. A. Kadhim, "Utilizing deep improved ResNet50 for brain tumor classification based MRI," *IEEE Open Journal of the Computer Society*, vol. 5, pp. 446–456, 2024, doi: 10.1109/OJCS.2024.3453924.

BIOGRAPHIES OF AUTHORS



Deepti Vadicherla    received the B.E. degree in computer science engineering from Shri Guru Gobind Singhji Institute of Engineering and Technology, India and M.E. in computer engineering from Savitribai Phule Pune University. She is the research scholar at G H Raisonni University, Amravati, India. She is the author of 4 publications in the field of machine learning. Her research interests include Machine learning, Data structure and Design and analysis of algorithms. She can be contacted at email: deepti.vadicherla@gmail.com.



Poonam Gupta    is Associate Professor at GH Raisonni College of Engineering and Management, India. She holds a Ph.D. degree in Computer Engineering. Her research areas are IoT, Data Mining, Computer Architecture and Databases. She is currently the Head of Department of Information Technology in GHRCEM. She has authored or coauthored more than 65 publications including 21 scopus indexed publications. She is a recipient of Edutech leadership award and most influential professors award of World education congress. Dr Gupta has two patents. Her research interests include Machine learning and Database. She can be contacted at email: poonam.gupta@raisonni.net.